BOIES SCHILLER FLEXNER LLP
David Boies (*pro hac vice*)
333 Main Street
Armonk, NY 10504
(914) 749-8200
dboies@bsfllp.com

Maxwell V. Pritt (SBN 253155)
Joshua I. Schiller (SBN 330653)
Joshua M. Stein (SBN 298856)
44 Montgomery Street, 41st Floor
San Francisco, CA 94104
(415) 293-6800
mpritt@bsfllp.com
jischiller@bsfllp.com
jstein@bsfllp.com

Jesse Panuccio (*pro hac vice*)
1401 New York Ave, NW
Washington, DC 20005
(202) 237-2727
jpanuccio@bsfllp.com

David L. Simons (*pro hac vice* forthcoming)
55 Hudson Yards, 20th Floor
New York, NY 10001
(914) 749-8200
dsimons@bsfllp.com

JOSEPH SAVERI LAW FIRM, LLP
Joseph R. Saveri (SBN 130064)
Cadio Zirpoli (SBN 179108)
Christopher K.L. Young (SBN 318371)
Holden Benon (SBN 325847)
Aaron Cera (SBN 351163)
Margaux Poueymirou (SBN 356000)
601 California Street, Suite 1505
San Francisco, California 94108
(415) 500-6800
jsaveri@saverilawfirm.com
czirpoli@saverilawfirm.com
cyoung@saverilawfirm.com
hbenon@saverilawfirm.com
acera@saverilawfirm.com

Matthew Butterick (SBN 250953)
1920 Hillhurst Avenue, #406
Los Angeles, CA 90027
(323) 968-2632
mb@buttericklaw.com

CAFFERTY CLOBES MERIWETHER & SPRENGEL LLP
Bryan L. Clobes (*pro hac vice*)
135 S. LaSalle Street, Suite 3210
Chicago, IL 60603
(312) 782-4880
bclobes@caffertyclobes.com

Counsel for Individual and Representative Plaintiffs and the Proposed Class.

UNITED STATES DISTRICT COURT
NORTHERN DISTRICT OF CALIFORNIA
SAN FRANCISCO DIVISION

RICHARD KADREY, et al.,

Individual and Representative Plaintiffs,

v.

META PLATFORMS, INC.,

Defendant.

CASE NO. 3:23-cv-03417-VC

DECLARATION OF DR. JONATHAN L. KREIN IN SUPPORT OF PLAINTIFFS' REPLY TO DEFENDANT META PLATFORMS, INC.'S OPPOSITION TO MOTION TO AMEND CASE MANAGEMENT SCHEDULE

I, Dr. Jonathan L. Krein, declare as follows:

1. I have personal knowledge of the matters stated herein and, if called upon, can competently testify thereto. I make this declaration pursuant to 28 U.S.C. § 1746 and Local Rule 6-3 in support of Plaintiffs' Reply to Defendant Meta Platforms, Inc.'s Opposition to Motion to Amend Case Management Schedule.

2. I am a data scientist and software engineer with broad expertise in machine learning and language models, artificial intelligence, and software engineering. I was retained by Plaintiffs in this case to provide expert testimony primarily predicated on my inspection of Meta's source code associated with its Llama Large Language Models ("Llama Models") in relation to data collection and processing, model training and fine tuning, prompt engineering, and the creation of any scripts or other components used to prevent the models from regurgitating copyrighted material. My *curriculum vitae* is attached as **Exhibit A** to this Declaration.

3. I have spent the last four days inspecting the source code Meta has made available to Plaintiffs. My understanding of Meta's source code has been further augmented by earlier reviews of this source code that were conducted by my colleague Don Langdon (on August 12-14, 2024, and September 27, 2024) and by my review of discovery, articles that Meta has published on its Llama Models, and general expertise in source code inspection.

4. As discussed below, I believe that Meta likely has not made available to Plaintiffs all relevant source code associated with its Llama Models—including pull requests reflecting interactions between developers and source code repositories, commits reflecting modifications to source code repositories (*i.e.*, code changes), and additional source code repositories not yet produced. Meta also produced additional source code, pull requests, and commits on October 2, 2024, that will require additional time to review and digest.

5. To date, Meta has produced three source code repositories:

a. [redacted]

b. [redacted]

c. [redacted]

6. Prior to October 2, 2024, Meta produced source code in each of the above repositories primarily in relation to its Llama 1 and Llama 2 models. There was also some source code associated with Llama 3. Because I do not know the full extent of Meta's source code with respect to these models, I cannot say whether Meta produced all source code that existed on the date Meta first collected this data (on or about April 4, 2024[1]). However, thus far I have been unable to find, for example, any source code related to how Meta trains its models to identify copyrighted material in order to prevent its models from regurgitating that material (or otherwise accomplishes the same, such as through tool use).

7. On October 2, 2024, I was told by Plaintiffs' counsel that Meta uploaded pull requests and issues to the source code inspection machine. A "**pull request**" is a mechanism to share source code changes and get feedback from other developers about these changes before adding them to the main codebase. An "**issue**" is a representation (*e.g.*, description) of a problem or task needing attention, the solution to which would typically involve a pull request being submitted along with associated source code changes for contribution to the main codebase. Neither a pull request nor an issue is source code. Both are producible as text files.

8. Prior to this date, I had not seen any pull requests or issues. On October 3, 2024, I again confirmed that no pull requests or issues were found anywhere among the materials Meta previously produced or made available for inspection. Further, when I was told that additional materials had been added to the source code inspection machine, I reviewed every volume that was available on the machine and determined that the new discovery still had not been made available.

[1] Based on the latest date found among the commits to the main branches of source code repositories Meta produced, as well as the timestamps of the files copied onto the source code inspection machine.

9. After Plaintiffs' counsel emailed one of Meta's lawyers about the missing discovery, someone from IT at Cooley entered the source code inspection room at approximately 2:45 p.m. local time without notifying me and accessed the source code inspection machine on which I had work product currently displayed. Once I discovered the IT personnel in the source code room and inquired, I was told that he entered to check something out on the source code inspection machine. However, he was not clear as to what he was actually there to do or had done—but once he left, I found an entirely new volume that had been made accessible within the source code inspection machine. I am certain the new volume was not present before the IT personnel arrived, as I had checked the list of available volumes earlier that day. The new volume included new source code in the form of source code repositories (including commits), as well as pull requests and possibly issues.

10. A source code "**commit**" is a record of a change made to a codebase within a source code repository, including, in this case, information on the changes themselves (*e.g.*, file names and line numbers changed, along with copies of the affected lines), as well as metadata about the changes such as author information, timestamps, and developer comments, among other things.

11. Specifically, Meta added new source code that likely had been created between April and September 2024 (along with the pull requests and issues).

12. Across the three source code repositories, which previously entailed about 29,000 commits, Meta added more than 26,000 new commits, nearly doubling the number of commits it previously made available for inspection. 26,000 new commits is a substantial volume of data that will require significant time to analyze, especially given the fact that, on average, these commits comprise numerous files each.

13. Also of particular significance, Meta made available for inspection for the first time on October 3, 2024, over 9,000 pull requests. [redacted] The pull requests show [redacted] being downloaded and processed. The pull requests also show emails between developers regarding the

work they were performing, along with thousands of descriptions of various tasks pertaining to the Llama Models. The pull requests represent a large portion of new and, in my opinion, highly relevant data that will require significant time to properly analyze.

14. In addition to needing time to review this new discovery, I also believe that Meta may not have produced all of the relevant pull requests for the Llama Models in its possession.

15. [redacted] However, this repository includes source code dating back to October 2, 2021. Typically, pull requests would be produced periodically and contemporaneously with the source code. [redacted]

16. [redacted] It is unclear why this repository ends in 2023, suggesting that additional source code may still need to be produced from this repository.

17. Regarding missing source code repositories, the source code that has been produced to date appears to be exclusively science and engineering code and no production or application code. "**Production**" code is the code that customers and users actually use. "**Application**" code is the code comprising a runnable computer program or system. Thus far, and with high likelihood, my review indicates that none of the production model code has been produced, including the application code that would encapsulate and/or exercise the production model code. As an example, for the meta.ai website to function, there must be APIs that facilitate access between the user interface and the production llama model(s) supporting the functionality available at that site. There must also be user interface code, among many other components. None of this code is found

among the source code so far produced.

18. The production model and application source code is relevant for at least three reasons: (i) it is the code defining the system that customers actually use; (ii) it is organized into a connected and complete flow, whereas the science and engineering code is, as typically expected, found in numerous disjoint pieces; and (iii) it likely contains relevant components that may not be found in the science and engineering code.

I declare under penalty of perjury that the foregoing is true and correct. Executed on the 4th day of October 2024 in San Francisco, California.

By: _____
Dr. Jonathan L. Krein

EXHIBIT A

Curriculum Vitae

Jonathan L. Krein

+1 843 861 3410 /m
jonathankrein@gmail.com

2155 E Blue Sky Dr
Eagle Mountain, UT 84005

Education

Doctor of Philosophy (Ph.D.), **Computer Science**, 2012–2014
Brigham Young University, Provo, Utah, USA

- Dissertation: *Replication and Knowledge Production in Empirical Software Engineering Research*
- Advisor: Charles D. Knutson
- Emphases: Software Engineering, Sociology, Bayesian Statistics, Machine Learning
- GPA: 4.00/4.00

Master of Science (M.S.), **Computer Science**, 2009–2011
Brigham Young University, Provo, Utah, USA

- Thesis: *Programming Language Fragmentation and Developer Productivity: An Empirical Study*
- Advisor: Charles D. Knutson
- Emphases: Software Engineering, Statistics, Data Mining
- GPA: 4.00/4.00

Bachelor of Science (B.S.), **Computer Science**, 2005–2008
Brigham Young University, Provo, Utah, USA

- GPA: 3.99/4.00, Summa cum Laude
- Clemson University, 2001–2002

Selected Courses: advanced topics in data mining, artificial intelligence, Bayesian methods in computer science, classical social theory, computer communications and networking, ethnographic research techniques, machine learning and data mining, research in open source systems, research in social media, social network analysis, software engineering, statistical methods for research, theoretical foundations of computing.

Positions & Affiliations

Partner, Chief Scientist - Data Science, Software Engineering (October 2020 – Present)
Crimson Vista (crimsonvista.com), Austin, Texas

Crimson Vista is a software and cybersecurity engineering firm focused on developing research and technology for industry applications (incl. government, academic, and non-profit sectors), as well as providing consulting services in both industry and litigation contexts. Industry services include engineering innovation, solutions engineering, software architecture and cloud systems engineering, cybersecurity planning and infrastructure development, software quality assurance and testing, compliance and best-practice assessments, security training, and others. IP litigation support includes experienced consulting and testifying subject matter experts skilled in source code review, systems analysis, compliance and best-practice assessments, prior art search and analysis, application of industry standards and best practices, etc. Research initiatives include P2PCD (a security certificate distribution protocol for vehicle-to-vehicle communication), CryptoDoneRight (a

cryptographic knowledgebase focused on practical cryptography and best practices), Network Security Playground (an instructional framework for teaching network security fundamentals), and privacy-preserving contact tracing for COVID-19 (and future pandemics).

Founder, Managing Partner (February 2020 – Present)
Dryden Technology Group (drydentechgroup.com), Draper, Utah

Dryden Technology Group is a boutique consulting firm providing support for intellectual property litigation (particularly trade secret, copyright infringement, contract dispute, and patent infringement cases) in the areas of software, source code, and related technologies. Dryden also provides a range of consulting services within the software industry, including in the areas of software development, software engineering, artificial intelligence, machine learning, data analytics, e-commerce, mobile application development, networking, web technologies, and others. Dryden works in partnership with other consulting firms to provide an expanded network of resources to its clients, including access to subject matter experts in a wide array of technologies from computer security to wireless data communications and more.

Selected Industry Engagements:

Data Scientist & Systems Architect (January 2021 – June 2023)
Kids On The Move (kotm.org), Orem, Utah

Kids On The Move (KOTM) is a non-profit serving children and families across Utah and Utah County – a region with a rapidly growing population (expected to double over the next 20-30 years). With client loads and waitlists increasing by double-digit percentages year over year, KTOM is overhauling its infrastructure for growth. I advise the initiative to build out KOTM's business intelligence and monitoring (data, systems, compliance, etc.) infrastructures – including advising the selection of tools and technologies, advising on the design of the system architecture, and helping develop key metrics, dashboards, and indicators. This initiative spans the organization as a whole, addressing each of KOTM's programs and serving all levels of relevant stakeholders from staff to directors, external partners, and donors and funding agencies.

Co-Founder, Managing Partner (August 2019 – Present)
Source Code Discovery (sourcecodediscovery.com), Austin, Texas

Source Code Discovery (SCD) is a consulting firm specializing in the review and analysis of software systems and source code for intellectual property litigation, mergers and acquisitions, and various cyber-security needs. SCD's services include extensive technical subject matter expertise in software development, software engineering, artificial intelligence, machine learning, data analytics, cyber-security, cryptography, e-commerce, wireless data communications, networking, embedded systems, web technologies, and others. SCD's experts have performed code reviews on every major operating system platform of the past 25 years and are knowledgeable in a vast array of programming languages and software frameworks. SCD's experts also have extensive experience testifying and consulting in complex intellectual property litigation.

Member, Board of Directors (June 2016 – Present)
Chair, Board of Directors (January 2021 – January 2023)
Kids On The Move (kotm.org), Orem, Utah

Advise on technology issues, provide software development services, and assist the organization and C-suite in meeting KOTM's mission, including overseeing administration of programs, finances and fundraising, strategic direction and planning. For example, managed the transition and enhancement of the volunteer portal from Friday's Kids Respite (see below) to KOTM. Serve on the executive committee, direct board business, represent KOTM and the board externally.

Positions: Member of the Board (June 2016 – Present), Chair-elect (July 2020 – January 2021), Member of the Executive Committee (July 2020 – Present), Board Chair (January 2021 – January 2023), Immediate Past Chair (advisory role, January 2023 – Present).

KOTM is a Utah-based nonprofit focused on serving children and families. Mission: Empowering families, one kid at a time, by supporting the development of children and families. Services: Early Head Start, Early Intervention, Autism Center, Child Care Center, Respite Care, Child & Family Mental Health.

Member, Board of Directors (August 2015 – Present)
Kinpoint (kinpoint.com), Alpine, Utah

Kinpoint is a genealogy software company that began in 2007 as a student project in the Computer Science Department at Brigham Young University (originally dubbed "The 20-Minute Genealogist"). After a five-year incubation on campus, Kinpoint launched as a corporation in late 2012. Kinpoint's mission is to engage the 90% who care about family history, yet (because it's too hard or otherwise inaccessible) spend little or no time actually doing it. Kinpoint is committed to delivering a meaningful family history experience in as little as 60 seconds. In February 2017, Kinpoint released the Take a Name® app which reached 500,000 installs in its first 24 months.

Professional Experience

Program Manager & Software Architect (January 2020 – March 2023)
Equity Values Foundation, Alpine, Utah

Manage software development for iOS/Android application software. Also manage strategic partnerships, as well as financial and other concerns related to software offerings.

Partner, Consulting and Testifying Expert (January 2013 – January 2022)
Ironwood Experts (ironwoodexperts.com), Alpine, Utah

Ironwood Experts was a consulting firm providing support for software IP litigation (especially patent and trade secret cases), as well as software process consulting and training. We provided expert services to some of the world's leading technology companies including Apple, Google, Microsoft, Amazon, Comcast, Toyota, Cisco, Barnes and Noble, Expedia, Vizio and others. We provided expert witness services, as well as technical support including prior art search, claim charts, source code analysis, and technical document review. We served on cases in district courts, the International Trade Commission, the Federal Trade Commission, and USPTO reexaminations. Ironwood Experts also provided software engineering consulting and training particularly focused on software process, software product quality, user interface and product design, and Amazon Web Services. Ironwood Experts also provided custom software testing services. Positions: Consulting Expert (Jan 2013 – Jan 2022), Partner (Jul 2014 – Jan 2022), Testifying Expert (Jan 2015 – Jan 2022), Managing Partner (Mar 2016 – Aug 2017).

Selected Industry Engagements:

AWS/Cloud & Systems Architect (May 2018 – November 2018)
Sound Concepts (now *Verb Technology*) (verb.tech), American Fork, Utah

Mature software company (offering tools for CRM, lead generation, and video marketing) needing specialized assistance with cloud architecture. I reduced their monthly AWS costs by 25% through infrastructure tuning and pruning. Also worked on a project to enhance and scale out their systems to handle significantly larger loads.

Software Process Advisor & Virtual CTO (October 2018 – July 2019)
Quintric (quintric.com), Alpine, Utah

Startup in the area of finance promoting a "real-time institutional grade cryptocurrency and legal tender backed by gold and silver" – that did not realize it was also a software company. I trained management on software process and foundational technology concerns, as well as provided user interface and product design consulting.

Kinpoint (kinpoint.com), Alpine, Utah

See board position above for description of Kinpoint.

President, CTO (January 2018 – January 2019)

Led the development of the Take a Name® app (including iOS and Android platforms, UI design, and backend architecture), which released in February 2017, reaching 500,000 installs in the first 24 months. Guided the software and systems architecture for Take a Name, including managed Kinpoint's AWS infrastructure and actively developed both front-end and back-end systems as needed (Java, JavaScript, Objective-C, etc.). Managed strategic partnerships, including with FamilySearch. Oversaw finances, hiring, and other business needs. Co-designed the product and managed quality assurance during development.

President, CEO (March 2016 – January 2018)

Brought in as part of a process to recycle the corporate DNA and to reboot on a new product (Take a Name). Enlisted to drive quality home, to ensure scalable architectures, and to act as a bridge between founder/vision/strategy and implementation. See above description for additional detail.

Inventor (2008 – 2016)

Patent: Genealogy Context Preservation. Issued May 28, 2013. US 8,452,805, see below.

Helped develop the early technology from 2008 to 2012. Tacitly involved at incorporation and from 2012-2016.

Adjunct Professor (December 2014 – May 2017)
Brigham Young University, Computer Science Department (cs.byu.edu), Provo, Utah

Taught CS 428, Software Engineering. The course teaches students to apply software engineering principles pertaining to product life cycles, requirements, analysis, specification, design, coding, testing, project management, quality assurance, and configuration management.

Member, Board of Directors (August 2013 – June 2016)
Friday's Kids Respite (fridayskids.org), Orem, Utah

Advised on technology issues, provided software development services, and participated in oversight of program administration, finances, and planning. Managed the online presence and technical infrastructure, including overseeing development/maintenance of the organization's website and internal business tools. Integrated the organization's PHP-based volunteer portal (which I previously built for them in 2007 – HTML/CSS/JavaScript frontend, MySQL/PHP backend) into a WordPress-templated architecture.

Friday's Kids Respite was a non-profit aimed at strengthening families of children with special needs by providing quality respite care. In June 2016, Friday's Kids Respite joined with Kids On The Move, a larger non-profit serving children and families (see above).

Adjunct Researcher (January 2014 – May 2016)
Research Scientist (September 2008 – December 2014)
Brigham Young University, SEQuOIA Lab (sequoia.cs.byu.edu), Provo, Utah

Research topics: the philosophy, methods, and challenges of replication in empirical software engineering research; the effects of language fragmentation (working concurrently across multiple programming languages) on developer productivity; coordination and collaboration concerns for stakeholders in large software project ecosystems; open source development processes (both in isolation and in comparison to proprietary development). Founded and organized IEEE's RESER workshop on Replication in Empirical Software Engineering Research.

Co-Founder, Managing Partner (July 2015 – April 2016)
Ironwood Quality Experts (*Ironwood QX*), Alpine, Utah

Provided software engineering consulting focused on software process and software product quality, as well as custom software testing services.

Web Developer (October 2007 – August 2013)
Friday's Kids Respite (fridayskids.org), Orem, Utah

Designed and implemented the Friday's Kids website and volunteer portal (PHP/JavaScript/HTML/MySQL framework). The website and portal supported client engagement and volunteer signup/participation, including providing user accounts, Wordpress-like dynamic webpage creation/editing via graphical interfaces, management of volunteer dates/times/schedules, automated email correspondence, and statistical reporting. Frontend implemented in HTML/CSS/JavaScript; backend in PHP; MySQL database; MVC architecture.

Friday's Kids Respite was a non-profit aimed at strengthening families of children with special needs by providing quality respite care. In June 2016, Friday's Kids Respite joined with Kids On The Move, a larger non-profit serving children and families (see above).

Instructor (June 2012 – August 2012)
Brigham Young University, Computer Science Department (cs.byu.edu), Provo, Utah

Taught CS 142, Introduction to Programming. The course introduces students to programming fundamentals in the C++ language.

Research Scientist (Intern) (July 2010 – October 2010)
IBM T.J. Watson Research Center, Governance Science Research Group (www.research.ibm.com/labs/watson), Hawthorne, New York

Software process governance. Interviewed software project stakeholders across IBM to identify coordination and collaboration concerns that threaten large software production environments. As a grounded theory study, the work informed initial hypotheses for process improvement. Further studied the problem of private information—referring to Friedrich Hayek's work, *The Use of Knowledge in Society*, 1945—as well as the potential applicability of principles from *The Wisdom of Crowds* (Surowiecki, 2005) and prediction markets to the management of private information within software production organizations.

Software and Solutions Engineer (April 2007 – August 2008)
Brigham Young University, Office of Information Technology (it.byu.edu), Provo, Utah

Developed technical solutions and software applications to enable internal business processes. Responsibilities included: software developer, project manager, team director, customer interface, requirements elicitation and definition. Emphasis on integrating business processes across the organization.

Research Assistant (January 2007 – December 2007)
Brigham Young University, Computational Science Laboratory (csl.cs.byu.edu), Provo, Utah

Bioinformatics research, Phylogenetic Search Open-source Data Analysis (PSODA). Designed and implemented an interpreted language (PsodaScript) for the specification and execution of phylogenetic search and alignment algorithms. PsodaScript extends the NEXUS format; designed to be a competitive, open source alternative to PAUP*, the current industry standard. PsodaScript is backwards compatible with PAUP*, but adds advanced language constructs—e.g., looping structures, conditionals, functions, etc. Used Flex/Bison (in C++) to implement the PsodaScript grammar. Upon execution, the PsodaScript interpreter constructs a call (or object) graph; the graph structure represents the flow logic, nodes represent program constructs.

Systems Management Engineer (February 2006 – April 2007)
Brigham Young University, Office of Information Technology (it.byu.edu), Provo, Utah

Maintained the university's network-monitoring framework. Developed plugins for and installed/maintained enterprise network-monitoring systems, such as ManagedObject, Nagios (open source), and Fruity (open source). Implemented enhancements and customizations for Nagios and Fruity. Nagios is an industry standard infrastructure-monitoring application; Fruity is a web-based configuration tool for the Nagios system. Of note: enhanced Fruity to support a distributed Nagios architecture; developed custom monitoring solutions for various technologies, including the campus emergency phone system and various web applications and server platforms.

Missionary, Ukraine (October 2002 – November 2004)
The Church of Jesus Christ of Latter-day Saints (lds.org), Kiev/Odessa, Ukraine

Taught the gospel of Jesus Christ in the Russian language. Taught English classes and rendered other service. Missions for the Church are full time, unpaid.

Academic Service

Program Advisor (June 2020 – July 2020)
Utah Valley University, Computer Science Department (uvu.edu), Orem, UT

Advised the creation of a new major in Computational Data Science.
See https://uvu.edu/cs/data_science.html.
See https://uvu.edu/catalog/current/departments/computer-science/computational-data-science-bs/.

Program Advisory Committee (January 2020 – January 2021)
Stevens-Henager College, School of Technology (stevenshenager.edu/technology), Orem, UT

Advised faculty on industry needs and recommendations for curriculum.

Co-Founder/Organizer

2013 IEEE International Workshop on Replication in Empirical Software Engineering Research (RESER 2013), co-located with the *International Symposium on Empirical Software Engineering and Measurement* (ESEM 2013, Baltimore, Maryland, USA).

2011 IEEE International Workshop on Replication in Empirical Software Engineering Research (RESER 2011), co-located with the *International Symposium on Empirical Software Engineering and Measurement* (ESEM 2011, Banff, Alberta, Canada).

2010 IEEE International Workshop on Replication in Empirical Software Engineering Research (RESER 2010), co-located with the *International Conference on Software Engineering* (ICSE 2010, Cape Town, South Africa).

Committee Chair Positions

IEEE International Workshop on Replication in Empirical Software Engineering Research (RESER 2013)
IEEE International Workshop on Replication in Empirical Software Engineering Research (RESER 2011)
IEEE International Workshop on Replication in Empirical Software Engineering Research (RESER 2010)

Reviewer/Program Committee

IEEE Transactions on Software Engineering (TSE), 2012–2018
Empirical Software Engineering: An International Journal (EMSE), 2012–2018
EMSE Special Issue on Experimental Replications, 2012, 2013
Working Conference on Mining Software Repositories (MSR), Data Track, 2013, 2016

External Reviewer

International Conference on Evaluation and Assessment in Software Engineering (EASE), 2013
IEEE Transactions on Software Engineering (TSE), 2010–2012
International Conference on Open Source Systems (OSS), 2011–2012
Journal of Information and Software Technology (IST), 2011
Psychology of Programming Interest Group Conference (PPIG), 2009–2011

Publications

(Electronic copies of papers available upon request.)

Peer-Reviewed Journals

1. **Jonathan L. Krein**, Lutz Prechelt, Natalia Juristo, Kevin Seppi, Aziz Nanthaamornphong, Jeffrey C. Carver, Sira Vegas, and Charles D. Knutson. A Method for Generalizing across Contexts in Software Engineering Experiments. *In process*.

2. **Jonathan L. Krein**, Lutz Prechelt, Natalia Juristo, Aziz Nanthaamornphong, Jeffrey C. Carver, Sira Vegas, Charles D. Knutson, Kevin D. Seppi, and Dennis L. Eggett. A Multi-Site Joint Replication of a Design Patterns Experiment using Moderator Variables to Generalize across Contexts. *IEEE Transactions on Software Engineering*, vol. 42, no. 4, pp. 302–321, 2016.

3. **Jonathan L. Krein**, Alexander C. MacLean, Charles D. Knutson, Daniel P. Delorey, and Dennis L. Eggett. Impact of Programming Language Fragmentation on Developer Productivity: A SourceForge Empirical Study. *International Journal of Open Source Software and Processes*, vol. 2, no. 2, pp. 41–61, 2010.

4. Hyrum D. Carroll, Adam R. Teichert, **Jonathan L. Krein**, Kenneth Sundberg, Quinn O. Snell, and Mark J. Clement. An Open Source Phylogenetic Search and Alignment Package. *International Journal of Bioinformatics Research and Applications*, vol. 5, no. 3, pp. 349–364, 2009.

Peer-Reviewed Conferences

5. Devin Robert Wright, Tim Severance, Charles D. Knutson, **Jonathan L. Krein**, Tyler D. Buchanan. An Autonomous Discord Bot to Improve Online Course Experience and Engagement: Lessons Learned Amid the COVID-19 Pandemic. *Proceedings of the IEEE Conference on Software Engineering Education and Training*, Maui, Hawaii, January 4-7, 2022.

6. (Awarded Best Paper) Brandon Foushee, **Jonathan L. Krein**, Justin Wu, Randy Buck, Charles D. Knutson, Landon J. Pratt, and Alexander C. MacLean. Reflexivity, Raymond, and the Success of Open Source Software Development: A SourceForge Empirical Study. In *Proceedings of the International Conference on Evaluation and Assessment in Software Engineering*, pp. 246–251, Porto de Galinhas, Brazil, 2013.

7. Quinn C. Taylor, **Jonathan L. Krein**, Alexander C. MacLean, and Charles D. Knutson. An Analysis of Author Contribution Patterns in Eclipse Foundation Project Source Code. In *Proceedings of the International Conference on Open Source Systems*, pp. 269–281, Salvador, Brazil, 2011.

8. **Jonathan L. Krein**, Patrick Wagstrom, Stanley M. Sutton Jr., Clay Williams, Charles D. Knutson. The Problem of Private Information in Large Software Organizations. In *Proceedings of the International Conference on Software and Systems Process*, pp. 218–222, Honolulu, Hawaii, USA, 2011.

9. Jason R. Casebolt, **Jonathan L. Krein**, Alexander C. MacLean, Charles D. Knutson, and Daniel P. Delorey. Author Entropy vs. File Size in the GNOME Suite of Applications. In *Proceedings of the International Working Conference on Mining Software Repositories*, pp. 91–94, Vancouver, Canada, 2009.

10. **Jonathan L. Krein**, Adam R. Teichert, Hyrum D. Carroll, Mark J. Clement, and Quinn O. Snell. PsodaScript: Applying Advanced Language Constructs to Open-source Phylogenetic Search. In *Proceedings of the Biotechnology and Bioinformatics Symposium*, pp. 89–94, Boulder, Colorado, USA, 2007.

Peer-Reviewed Workshops

11. Kyle L. Blatter, T.J. Gedhill, **Jonathan L. Krein**, and Charles D. Knutson. Impact of Communication Structure on System Design: Towards a Controlled Test of Conway's Law. In *Proceedings of the International Workshop on Replication in Empirical Software Engineering Research*, pp. 25–33, Baltimore, Maryland, USA, 2013.

12. Sabrina E. Bailey, Sneha S. Godbole, Charles D. Knutson, and **Jonathan L. Krein**. A Decade of Conway's Law: A Literature Review from 2003–2012. In *Proceedings of the International Workshop on Replication in Empirical Software Engineering Research*, pp. 1–14, Baltimore, Maryland, USA, 2013.

13. **Jonathan L. Krein**, Landon J. Pratt, Alan B. Swenson, Alexander C. MacLean, Charles D. Knutson, and Dennis L. Eggett. Design Patterns in Software Maintenance: An Experiment Replication at Brigham Young University. In *Proceedings of the International Workshop on Replication in Empirical Software Engineering Research*, pp. 25–34, Banff, Alberta, Canada, 2011.

14. Scott H. Burton, Paul M. Bodily, Richard G. Morris, Charles D. Knutson, and **Jonathan L. Krein**. Design Team Perception of Development Team Composition: Implications for Conway's Law. In *Proceedings of the International Workshop on Replication in Empirical Software Engineering Research*, pp. 52–60, Banff, Alberta, Canada, 2011.

15. **Jonathan L. Krein** and Charles D. Knutson. A Case for Replication: Synthesizing Research Methodologies in Software Engineering. In *Proceedings of the International Workshop on Replication in Empirical Software Engineering Research,* Cape Town, South Africa, 2010.

16. Alexander C. MacLean, Landon J. Pratt, **Jonathan L. Krein**, and Charles D. Knutson. Threats to Validity in Analysis of Language Fragmentation on SourceForge Data. In *Proceedings of the International Workshop on Replication in Empirical Software Engineering Research*, Cape Town, South Africa, 2010.

17. Alexander C. MacLean, Landon J. Pratt, **Jonathan L. Krein**, and Charles D. Knutson. Trends that Affect Temporal Analysis Using SourceForge Data. In *Proceedings of the International Workshop on Public Data about Software Development*, South Bend, Indiana, USA, 2010.

18. **Jonathan L. Krein**, Alexander C. MacLean, Daniel P. Delorey, Charles D. Knutson, and Dennis L. Eggett. Language Entropy: A Metric for Characterization of Author Programming Language Distribution. In *Proceedings of the International Workshop on Public Data about Software Development*, Skovde, Sweden, 2009.

19. Charles D. Knutson and **Jonathan L. Krein**. The 20-Minute Genealogist: A Context-Preservation Metaphor for Assisted Family History Research. In *Proceedings of the Workshop on Technology for Family History and Genealogical Research*, Provo, Utah, USA, 2009.

Thesis and Dissertation

20. **Jonathan L. Krein**, *Replication and Knowledge Production in Empirical Software Engineering Research*. Doctoral Dissertation, Brigham Young University, 2014.

21. **Jonathan L. Krein**. *Programming Language Fragmentation and Developer Productivity: An Empirical Study*. Master's Thesis, Brigham Young University, 2011.

Trade Journals

22. **Jonathan L. Krein**. RAM Nagios vs. HD Nagios: Performance Evaluation. *Sys Admin, the journal for UNIX and Linux systems administrators*, vol. 16, no. 3, pp. 32–34, 2007.

23. **Jonathan L. Krein**. Nagios and Fruity: What is Their Monitoring Potential for Your Network? *Sys Admin, the journal for UNIX and Linux systems administrators*, vol. 16, no. 1, pp. 12–16, 2007.

Patents

24. United States Patent No. 8,452,805. Charles D. Knutson, **Jonathan L. Krein**, Daniel Zappala, and Daniel P. Delorey. *Genealogy Context Preservation*. Issued: May 28, 2013.

Other Publications

25. **Jonathan L. Krein**, Charles D. Knutson, and Christian Bird. Report from the 3rd International Workshop on Replication in Empirical Software Engineering Research (RESER 2013). *SIGSOFT Software Engineering Notes*, vol. 39, no. 1, pp. 31–35, 2014.

26. **Jonathan L. Krein**, Charles D. Knutson, Lutz Prechelt, and Christian Bird. Message from the RESER 2013 Workshop Chairs. In *Proceedings of the International Symposium on Empirical Software Engineering and Measurement*, p. 395, Baltimore, Maryland, USA, 2013.

27. **Jonathan L. Krein**, Charles D. Knutson, Lutz Prechelt, and Natalia Juristo. Report from the 2nd International Workshop on Replication in Empirical Software Engineering Research (RESER 2011). *SIGSOFT Software Engineering Notes*, vol. 37, no. 1, pp. 27–30, 2012.

28. Charles D. Knutson, **Jonathan L. Krein**, Lutz Prechelt, and Natalia Juristo. Report from the 1st International Workshop on Replication in Empirical Software Engineering Research (RESER 2010). *SIGSOFT Software Engineering Notes*, vol. 35, no. 5, pp. 42–44, 2010.

29. Charles D. Knutson, **Jonathan L. Krein**, Lutz Prechelt, and Natalia Juristo. 1st International Workshop on Replication in Empirical Software Engineering Research (RESER). In *Proceedings of the International Conference on Software Engineering*, pp. 461–462, Cape Town, South Africa, 2010.

30. Charles D. Knutson and **Jonathan L. Krein**. The 20-Minute Genealogist: Validating a Research Metaphor. Poster in *BYU Studies 50th Anniversary Symposium*, Provo, Utah, USA, 2010.

Honors and Awards

- Recipient:
  - Graduate Research Fellowship Award, Brigham Young University, 2010
  - Karl G. Maeser Scholarship, full tuition and books, Brigham Young University, 2007
  - Brigham Young Scholarship, full tuition, Brigham Young University, 2005–2006, 2008
  - Palmetto Fellows Scholarship, full tuition, Clemson University, 2001–2002
- Graduated Summa cum Laude, B.S. in Computer Science, Brigham Young University, 2008
- Dean's List all years, Clemson University and Brigham Young University
- Honor Society of Phi Kappa Phi, Brigham Young University chapter, invited

Affiliations

- Institute of Electrical and Electronics Engineers (IEEE), *Member*, 2010 – 2015
- Association for Computing Machinery (ACM), *Member*, approx. 2010 – 2015

Miscellaneous

- **Citizenship:** United States of America.
- **Spoken languages:** Native English, Russian (lived two years in Ukraine, 2002–2004).
- **Programming languages, frameworks, tools, etc.:** Java, C/C++, Objective-C, C#, PHP, Python, Perl, JavaScript, CSS, HTML/HTML5, COBOL, R, SAS, Node.js, MySQL, PostgreSQL, iOS, Android, Unix, Linux, macOS, Windows, AWS, Apache (HTTP Server), Maven, Make, Oracle, Flex/Bison, Hibernate, Tapestry, JSP, JSF, Angular, Ruby, Scheme, JBoss, LaTeX, etc.
- **Statistical methods/tools:** Linear mixed models, Bayesian inference, SAS, R, etc.
- **Empirical research methods:** Grounded theory, ethnographic research, surveys and interviews, human subjects research, study design, IRB processes, subject recruitment, data cleaning/analysis, etc.
- **Data science:** Machine learning, data mining, social network analysis.
- **Other methods:** research in open-source software, analysis of software repositories, source code analysis.